

Noname manuscript No.
(will be inserted by the editor)

Autocalibration based partitioning relationship and parallax relation for head mounted eye trackers

Sacha Bernet · Christophe Cudel · Damien Lefloch · Michel Basset

Received: date / Accepted: date

Abstract This paper presents new methods to calibrate a head mounted Eye Tracker (ET) automatically, as well as a new way to obtain an estimated Point Of Regard (POR), taking account of the parallax. Calibration is performed in real time; it is easy for the user who just needs to look at one calibration pattern for a few seconds before starting. This method provides a very important couple of points which help to use a local relationship to compute the POR instead of a global one. This approach significantly improves the precision of the points of regard when the scene camera is mounted with a short focal lens. An estimation of POR when the user looks somewhere outside the calibration distance is also proposed. This estimation is based on an ET modelling such as a stereovision system, to take account of the parallax effect. The aim of this study is to simplify the use of ET techniques for "non-initiated" people, especially here learner drivers.

Keywords Eye tracking · calibration · image processing

1 Introduction

Eye Trackers (ET) systems are increasingly used in different fields such as ophthalmology, cognitive science and communication by gaze interaction. Corresponding applications are modelling tools for vehicle driver

Sacha Bernet
MIPS Université de Haute Alsace
12, rue des frères Lumière
68093 Mulhouse cedex
Tel.: +33 3 89 33 69 51
Fax: +33 3 89 42 32 82
E-mail: {sacha.bernet, christophe.cudel, damien.lefloch, michel.basset}@uha.fr

behaviour [1], young children [13] or web page reading [6], and aid to disabled people [23], video games [17], flying simulators [27] and Human Computer Interface.

An eye tracker system helps to determine in which direction a person looks. There are two families [10] of eye tracker systems. The first one is composed of one or two cameras integrated in flat screen base and facing at the user and his eyes. The advantage of remote systems is that the user just needs to watch the scene without worrying about the vision system, but it forces the user to always face the cameras to allow gaze detection. The second family consists of two cameras: the first one is the "scene camera" and produces images of the user's field of view; the second one is the "eye camera" which observes the users eye, and it is used to extract the pupil position. Both cameras are carried by the user, which helps to obtain the gaze direction or the Point Of Regard (POR) in any head position.

Any eye tracker system needs a learning stage called calibration before being used. The calibration calculates a relationship between the eye pupil positions and their corresponding points in scene images. The quality of this first phase is crucial for the accuracy of measurements, and it must be as simple as possible for the user. In general, the user successively looks at series of points arranged on a calibration grid. Following the ET family used, a grid of points is displayed on a screen for remote ET, or placed in front of the user for a head mounted ET.

This paper proposes a new approach to the calibration of ET systems, using a calibration with a single point to simplify the learning phase. In experiments, this point is characterized by a target described in the next section.

An interesting consequence of this approach is that it provides greater precision in determining the relationship to compute the POR. ET modelling as a stereovision system also shows that it is possible to estimate the POR more accurately when the user looks at something outside the distance used for calibration.

The first section will describe the hardware and image processing chosen due to specific constraints of the project. The second part presents a calibration model, and the third one an auto-calibration method which uses only one target. This new approach is validated by simulations and experimental setups. In the fourth section, the POR is computed more accurately using a local relationship instead of a global one. The local relation is only possible if it is coupled with the proposed auto-calibration method due to an important amount of correspondence points. A new contribution is proposed in the fifth section in order to improve the POR for situations where the gaze of the user is outside the range of the calibration distance. A parallax error is computed using a ET stereovision modelling. This best estimation is the result of a parallax error computed with a stereovision modelling of the ET. The last section includes a discussion of future works and a conclusion.

2 The eye tracking system

This part describes the eye tracker system developed in the laboratory [Fig. 1] for learner drivers within the TRACECARD¹ project and explains the choice of the technology hardware and basic algorithms used for pupil and target detection. Since the POR of the driver is relevant in any head position, a headband eye tracker system was chosen and designed instead of a remote ET.

The ET system is mounted on a headband where the cameras are attached. The user adjusts the ET position on his head until his eye is well centered in the image of the eye camera. Cameras are linked together and connected by a single cable to the computer. Concerning this real time application, the algorithms for calibration and detection are designed so that the user does not have to adjust anything. Pupil detection and the calibration target processes will be described in the next section.

¹ TRACECARD, Training Capabilities for Efficient CAR Driving, is a project funded by a European programme, to propose innovative training tools for learning to drive.

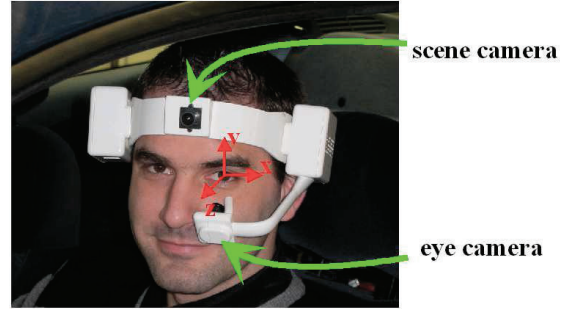


Fig. 1 Head mounted eye tracker system developed in the MIPS laboratory

2.1 Pupil detection

To meet the strong variation of brightness encountered in real situations an algorithm for robust pupil detection is necessary. The user's eye is illuminated by an infrared (IR) led [18] [29] and the lens of the eye camera is equipped with an IR pass filter. This IR illumination provides constant brightness on the eye. IR led illumination is advantageously used in numerous ET systems, generally coupled with the corneal reflection (CR) (first Purkinje image [28] and [20]) techniques [Fig. 2 a]. It is difficult to use CR in an outdoor environment where multiple reflection can be found in the eye image as shown in [Fig. 2 b]. The centre of the pupil is discussed in this paper since it provides for our specific application the best results. Note that our approach does not restrict the use of CR information.

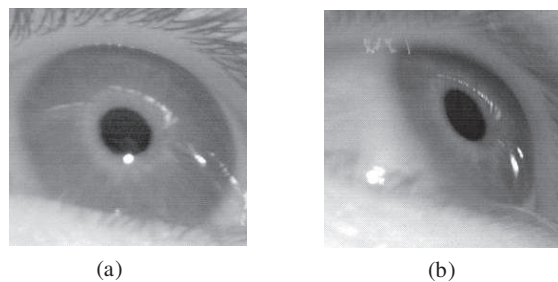


Fig. 2 Reflection of the infrared source on the eye. (A) corneal reflection can be exploited in this case, (b) multiple spotlights make corneal reflection difficult to exploit.

There are numerous algorithms for pupil centre detection in the literature [19] [8] [32] [9]. Most of them need the adjustment of some parameters or thresholds. This is uncomfortable for people who are not familiar with ET systems. Automatic methods with no settings, such as the starburst algorithm for example [8] are preferred here.

The proposed algorithm uses ray tracing to find the points belonging to the ellipse which best describes the pupil. The different stages are (see [Fig. 3]):

1. Initially the software performs an "integral image" [4] on the eye image.
2. from an integral image algorithm, the coordinates of a point (P_{ii}) are detected. This point represents the region where the sum of the intensity pixels is minimum. This point is located in the pupil area;
3. the ray is computed in all directions from P_{ii} ;
4. on each ray, the first gradient with a significant value is chosen;
5. with the set of chosen points, the best ellipse is fitted using a RANSAC algorithm [12];
6. finally, the centroid of the ellipse is calculated and chosen as the centre of the pupil.

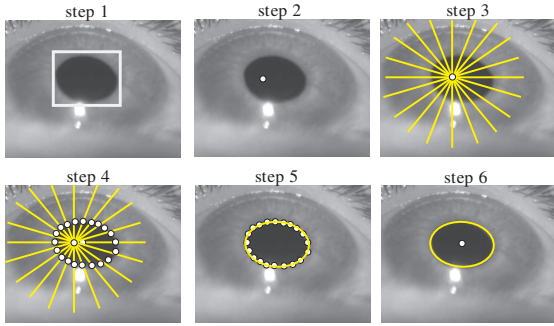


Fig. 3 The different steps of the algorithm for detection of the pupil centre developed in our laboratory.

2.2 Target detection

A special target has been designed for automatic detection in the scene image and is used in the auto-calibration process proposed in this paper. The algorithm can detect the target whatever the camera position, and is robust to rotations and perspective transformations. The pattern is composed of two black squares nested one inside the other [Fig. 4]. Following the calibration distance, the target can take different sizes, but the ratio between the squares remains the same.

The algorithm for pattern detection applied on images of the scene camera during the calibration phase is decomposed as follows:

1. binarization of greyscale images with a standard adaptive threshold method;
2. contour extraction of binary patterns;

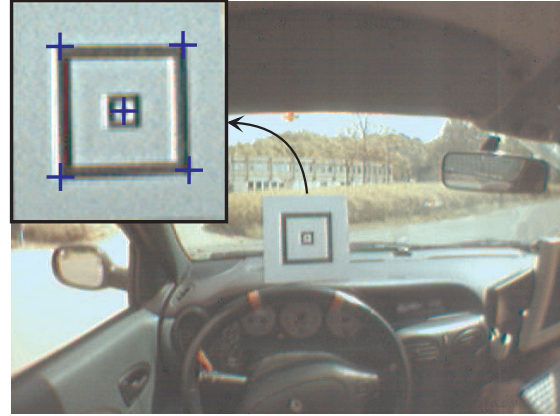


Fig. 4 Special pattern to simplify the calibration phase.

3. each contour is approximated as a polygon of 4 vertices;
4. for each external contour, 2 internal contours are searched;
5. when the ratios between these contours correspond to those of the target, the 4 corners of the external contours are used to compute the centre of the target;

This algorithm works for all outdoor illumination configurations tested thanks to its unique characteristic and high contrast difference

3 Eye tracker model and standard calibration

The calibration is an essential step for ET systems. It involves finding a mapping relationship which connects the centre of the pupil and a point viewed by the user. The standard calibration consists in collecting the pupil centre in eye camera and its corresponding points in the scene camera. Most of ET systems use a homography, a polynomial relation or a neural network as mapping expression [15, 26, 25, 24]. Polynomial expressions are a common solutions to relie 2D to 2D images. Apparently [30] in 1973 was the first to investigate the properties of polynomial system to map images. For Eye Tracking system, most of papers seem to accept that polynomial expression is the most suitable relationship for mapping. In [5], we have compared different kind of polynomial expression and concluded that a 3-order polynomial produces the best mapping. Standard calibration consists in collecting a set of pair points provide when the user looks at the calibration grid (usually based on 3x3 or 4x4 points). In general, the pair of points are composed of the pupil centre from the eye camera and its corresponding POR from the scene camera

(current point in the grid). To avoid parallax effects, the points have to be located in a plane, and the user must keep his head still during the entire process. The calibration scheme can differ depending on eye tracker hardware principle. For example [21] proposes a system with a set of mirrors and a single camera where the images contains the scene and eye projections. This system has been proposed for young children whose are not old enough to understand the calibration process. In this case, calibration is done off line, and correspondences are ensured with different coloured pattern to fix children attention and gaze on these patterns. With others systems, the calibration can be supervised. A recent commercial eye tracker [22], which uses some IR pattern for the calibration, proposes a supervised calibration by a software on an embedded system. The calibration process is then easier, however requires still the presence of a supervisor. In conclusion, collect the correspondences is not a friendly operation for inexperienced users and is difficult to perform alone. We propose in the next paragraphe to explain that once the mapping relationship is obtained, is possible to reverse the calibration scheme. In this case, the user needs simply to watch a single point while moving his head. The calibration procedure will be simplified and performed without external assistance.

3.1 Eye tracker model

The mapping relationship used for a monocular head mounted ET is a composition of homography and polynomial expressions. It is now necessary to study how a mapping relationship can connect the centre of the pupil and the point watched by the user.

There are several methods of calculation in order to find a relationship between the images of the eye and those of the scene. The most frequent ones in the literature are: homography and the polynomial relationship of order 1 up to 3. Homography [16] is a relationship passing from plane to plane, which requires images without distortion. Distortions can be corrected using Zhang method [33]. The polynomial relationship has the advantage to directly integrate optical distortions [14, 26, 25].

Figure [Fig. 5] shows a simplified view from above of the eye tracker system carried by the user. It is then possible to write the following relations for the projection of a point P in the scene camera, noted p_s with a homography:

$$p_s = AP \quad (1)$$

$$s \begin{pmatrix} u_{p_s} \\ v_{p_s} \\ 1 \end{pmatrix}_{(3,1)} = \begin{pmatrix} A_{11} & \dots & A_{13} \\ \vdots & \ddots & \vdots \\ A_{31} & \dots & A_{33} \end{pmatrix}_{(3,3)} \begin{pmatrix} X_P \\ Y_P \\ Z_P \end{pmatrix}_{(3,1)} \quad (2)$$

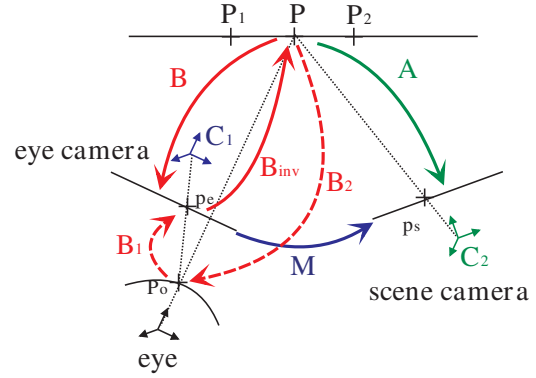


Fig. 5 Schematic top view of standard calibration. The projection of point P in the scene camera is p_s and the projection of the pupil centre looking at point P is p_e on the eye camera. P , P_1 and P_2 : points on the pattern.

C_1 and C_2 : optical centre of the cameras.

P_o : Intersection of visual axis with the cornea.

p_e : projection of point P_o in the eye camera.

p_s : projection of point P in the scene camera.

A and B : relation between P and cameras (p_s and p_e).

M : relation between the points p_e and p_s .

The projection of the pupil centre looking at point P on the eye camera can be decomposed as follows:

$$P_o = B_2P \text{ and } p_e = B_1P_o \quad (3)$$

$$p_e = B_1B_2P \quad (4)$$

$$p_e = BP \quad (5)$$

The eye camera and eye can be approximated as a system composed of a camera and a particular mirror causing deformation due to the particular shape of the eye. This deformation is close to an optical distortion effect and therefore the relationship between point P and point p_e can be expressed as a polynomial function:

$$\begin{pmatrix} u_{p_e} \\ v_{p_e} \\ 1 \end{pmatrix}_{(3,1)} = \begin{pmatrix} B_{11} & \dots & B_{1n} \\ \vdots & \ddots & \vdots \\ B_{31} & \dots & B_{3n} \end{pmatrix}_{(3,n)} \begin{pmatrix} X_P^m \\ Y_P^m \\ Z_P^m \\ \vdots \\ X_P \\ Y_P \\ Z_P \end{pmatrix}_{(n,1)} \quad (6)$$

Where n is the size of the matrix according to the polynomial degree. m is the order of the polynomial. The relationship B can go from point P to p_e , but it is not possible to write $P = B^{-1}p_e$ because B is not a square matrix. However, there is a relationship which goes from point p_e to point P .

$$P = B_{inv}p_e \quad (7)$$

$$\begin{pmatrix} X_P \\ Y_P \\ Z_P \end{pmatrix}_{(3,1)} = \begin{pmatrix} B_{inv11} & \dots & B_{inv1n} \\ \vdots & \ddots & \vdots \\ B_{inv31} & \dots & B_{inv3n} \end{pmatrix}_{(3,n)} \begin{pmatrix} u_{p_e}^m \\ v_{p_e}^m \\ \vdots \\ u_{p_e} \\ v_{p_e} \\ 1 \end{pmatrix}_{(n,1)} \quad (8)$$

$$B_{inv} \neq B^{-1} \quad (9)$$

Finally this model gives the relationship between the centre of the pupil p_e and the corresponding point p_s in the scene camera, using equations 2 and 8.

$$p_s = AB_{inv}p_e = Mp_e \quad (10)$$

3.2 Standard calibration

M is obtained during calibration performed before the eye tracking analysis. The standard method for calibration consists in placing between 9 and 16 points P in a plane located in the field of view of the user. The user has to successively look at these points without moving his head. For each acquisition, the pupil centre p_e is computed from the eye camera and the point looked

at by the user, p_s , is extracted from the scene camera. These corresponding points are used to compute the M relationship with a standard minimization method such as LMS or Levenberg-Marquad [16].

Standard calibration gives good results but is sometimes difficult to perform correctly for an inexperienced ET user. Three main reasons explain why it is not an easy operation. First, the user must not move his head during calibration, which can be difficult for non regular users such as the learner drivers. Secondly, in the automotive field, it is not very easy to find the best location for a grid of aligned points. Thirdly, if the calibration points are not automatically detected in the images of the scene camera, another person must help to select them manually. The method proposed in the next section solves these three points with a fast and real-time calibration which does not interfere its quality. The calibration scheme can be reversed to simplify the calibration procedure.

4 Auto-calibration

The proposed auto-calibration method consists in reversing the standard calibration scheme. Instead of looking at points located in a calibration plane, the user gazes at a single point and moves his head to obtain different calibration positions [Fig. 6]. In this case, the relationship M obtained is similar to the one with standard calibration, if the user moves his head in a vertical plane. In the next section, a calibration simulation for ET is described in order to compare the reliability of both approaches.

4.1 Simulations

Standard calibration and auto-calibration are compared. Simulations are carried out by using *ET-simul* [3] *MatLab* toolbox framework. This framework is dedicated to ET system simulation and it is based on a realistic optical eye model. This model takes into account the corneal and pupil shapes and the shift existing between the optical and visual axes.

Figure [Fig. 7] compares standard calibration with a target of 25 points spaced out 10 cm and apart auto-calibration with a single point. Both calibrations are performed at distance of two metres. The head remains fixed for the first one; for the second, simulated user continuously looks at the centre of the target and moves his head in x , y directions by steps of 10 cm. With the sets of correspondences obtained, the relationships for

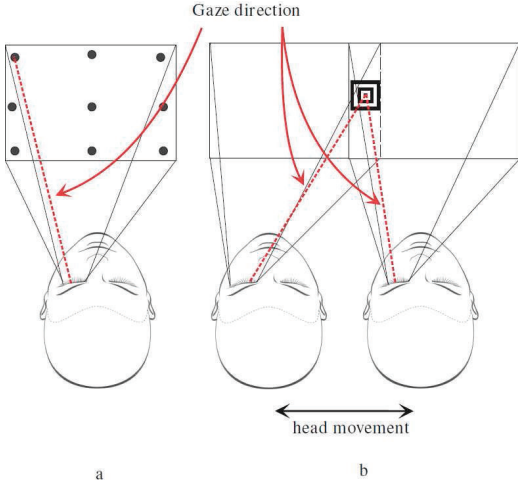


Fig. 6 Different methods of calibration for ET systems. a : standard calibration, the user does not move and successively looks each calibration point. b : auto-calibration, the user keeps his gaze on the pattern and moves his head at the same time.

both calibrations are computed with a third-order polynomial as described in equation 11. We have retained this relation, according to our previous results [5].

$$\begin{aligned}
 x_s &= a_{1.1} + a_{1.2}x_e + a_{1.3}y_e + a_{1.4}x_e y_e + a_{1.5}x_e^2 + \\
 &\quad a_{1.6}y_e^2 + a_{1.7}x_e^2 y_e + a_{1.8}x_e y_e^2 + a_{1.9}x_e^3 + a_{1.10}y_e^3 \\
 y_s &= a_{2.1} + a_{2.2}x_e + a_{2.3}y_e + a_{2.4}x_e y_e + a_{2.5}x_e^2 + \\
 &\quad a_{2.6}y_e^2 + a_{2.7}x_e^2 y_e + a_{2.8}x_e y_e^2 + a_{2.9}x_e^3 + a_{2.10}y_e^3
 \end{aligned}
 \tag{11}$$

with

x_s et y_s : calculated points in the scene image in pixels;

x_e et y_e : measured points in the eye image in pixels;

a_{xx} : coefficients to be determined during the calibration phase;

Figure [Fig. 7] shows the results of POR estimation after standard and auto-calibration. The POR computed with both methods are well aligned with the real points directly projected in a scene camera. Mismatch positions do not exceed 0.2 pixel and can be considered as rounding errors in the computation.

This simulation works perfectly in perfect calibration conditions. Now, in practice, the user has some difficulties to stay in the same vertical plane when he moves his head during calibration. In figure [Fig. 8] the previous simulations are reproduced by adding a random head movement in a z direction (± 5 cm maximum). In

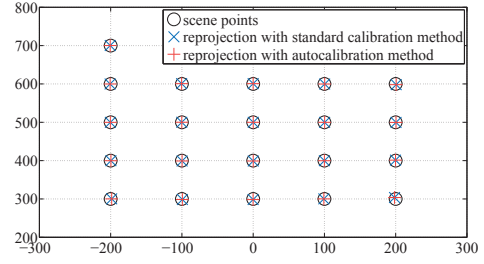


Fig. 7 Reprojection between different calibration methods. Standard calibration: the head remains fixed and the calibration points are in a plane. Auto-calibration with a single point: the point remains fixed but the head and the vision system move in a plane. Both calibrations are performed at a distance of 2m.

this case, the POR estimation is not perfectly exact, and a mean error less than 4 pixels is observed. As will be seen in part 6, it is possible to reposition the POR according to the working distance.

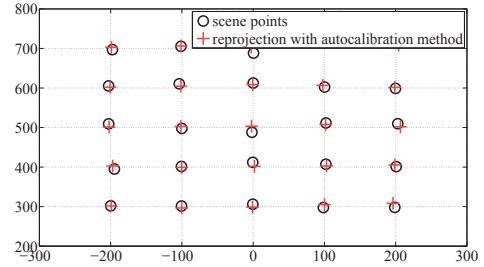


Fig. 8 Influence of head movement of plus or minus 5 cm in z during the calibration phase.

4.2 Experimentals results

In figure [Fig. 9] auto-calibration for ten seconds potentially gives 300 correspondences (acquisition at 30 frames per second). Some correspondences are mismatches, essentially when the user was distracted for a few milliseconds during calibration. These outliers can be rejected through a filtering with RANSAC algorithm [12]. To validate this new method of calibration experimental, several series of calibrations were performed : standard calibrations with fixed head and calibrations with the auto-calibration method proposed.

The experimental phase comparison of two calibrations methods is not possible in the same conditions as for simulations, because it is obvious that the same conditions cannot be recreated for each calibration [Fig. 9].

It is still possible to compare then with good accuracy. To do so several standard calibrations are implemented with 9 points of measurement; auto-calibration with a single point where mismatched correspondences are rejected with RANSAC algorithm. The user at 2 metres from the target and the target is automatically detected in both cases using the algorithm developed in the laboratory and presented above.

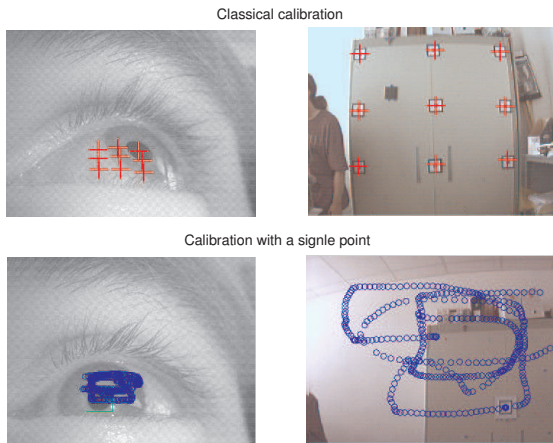


Fig. 9 Example of points selected for standard calibration and calibration with a single point. For calibration with a single point : the blue circles are any points of the calibration.

[Table. 1] shows the comparison between standard calibration and auto-calibration with a single point. It appears that the reprojection errors of the two methods are in the same order of magnitude. Either method gives quite acceptable and above all exploitable results. But the reprojection errors in both cases are more important on the image borders than in the centre. This can be explained by the short focal lens used in the experiment. As will be seen in the next part it is possible to reduce this error by taking advantage of auto-calibration.

5 Relationship based partitioning

The large number of correspondence points is used here to compute a local relationship. The aim is to increase the POR estimation, essentially on the borders of the scene image. If the global relationship with a third order polynomial presents a mean precision lower than 3 pixels [Table. 1], it can be observed that precision is not constant on the image surface and decreases on the image borders. There are two reasons to explain this effect. First, the spherical shape of the eye is not exactly taken into account by the third polynomial order relation used. Secondly, the focal lens mounted on

the scene camera can increase this effect, a short focal lens will produce more distortion, which is difficult to compensate. So a local relationship instead of a global one is proposed for the best compensation of these distortions.

The proposed method consists of keeping in a database the set of correct correspondences obtained during calibration. When ET analysis starts, for each pupil position, are collected in the database the closest correspondences and computed with a local relationship. Figure [Fig. 10] shows the principle of this local calibration method, decomposed into 4 steps:

1. the first step consists in detecting the pupil centre position in the eye image;
2. the second step is the selection in the calibration database (blue points) the closest pupil positions included in a radius of 30 pixels from the current pupil centre position (Positions included in the green circle). If there are no sufficient matches the radius is enlarged to include twenty correspondences;
3. the third step consists in selecting the matching scene correspondences in the database;
4. The last step is the calculation of a local relationship with the set of correspondences selected by using the same polynomial expression as global relationship (11).

5.1 Simulations

For the results presented in this section, the calibration distance is set to $2.5m$ and the scene camera is mounted with a $4.3mm$ focal lens. Optical distortion is not corrected on eye and scene images. Figure [Fig. 11] shows a simulation to compare auto-calibration with the global and the local relationships. Figure [Fig. 12] shows the same results when Gaussian noise is added to the coordinates points of the database calibration. In both cases, the reprojection error is small near the image centre, but this error increases for the points located on the border of the image. In these cases, the local relationship improves the results and reduces the distortion due to the spherical shape of the eye. This improvement is confirmed by the results of [Table. 2] where it clearly appears that the mean error and standard deviation, computed with the data of figures [Fig. 11] and [Fig. 12], is reduced when local relationships are used.

Figure [Fig. 13] simulates the mean reprojection error when optical distortion is added to the images of scene

method	Global relationship			
	Standard calibration		Auto-calibration	
	error in pixels	error in degrees	error in pixels	error in degrees
1 st order	15.6	1.0	6.2	0.4
2 nd order	6.9	0.4	3.9	0.3
3 rd order	3.9	0.3	2.7	0.2

Table 1 Comparative table between standard calibration and calibration with a single point. The distortion is 0.8%

measurement	auto-calibration without noise	
	global relationship	local relationship
mean error in pixels	8.18	2.22
standard deviation in pixels	6.38	1.42
measurement	auto-calibration with noise	
	global relationship	local relationship
mean error in pixels	9.56	3.94
standard deviation in pixels	6.80	2.54

Table 2 Comparative table between auto-calibration without noise and auto-calibration with noise.

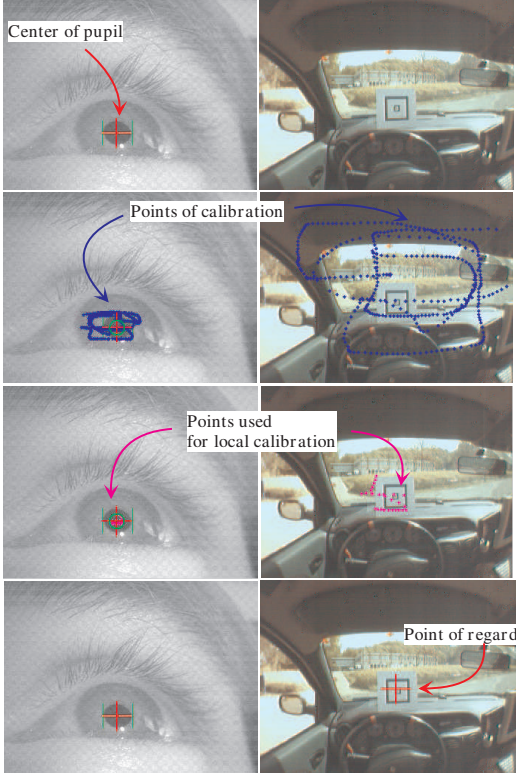


Fig. 10 Calculation of the relationship of local crossing through the points which belong to the green circle.

camera. Here, the idea is to simulate the benefits of the local relationship when short focal lenses are used for the scene camera. The distortion is given in percent, and computed by these standard definitions:

$$x_{dist} = x + (x - x_c) * (k_1 * r^2 + k_2 * r^4) \quad (12)$$

$$y_{dist} = y + (y - y_c) * (k_1 * r^2 + k_2 * r^4) \quad (13)$$

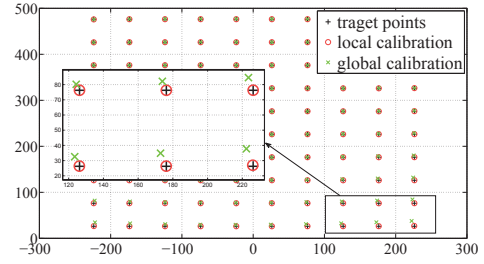


Fig. 11 Simulation without noise on the data to compare the calculation of a global and local relationship. The calculation with a local relationship cancels the effects of edges.

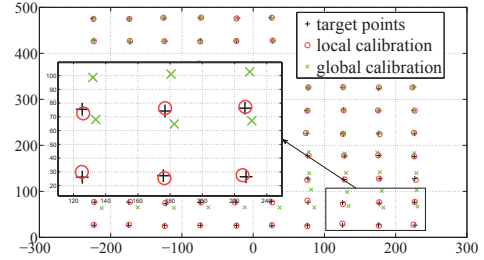


Fig. 12 Simulation with noise on the data to compare the calculation of a global and local relationship. The calculation with a local relationship improves the reprojection considerably.

with

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2} \quad (14)$$

x_{dist} and y_{dist} : distorted points;

x and y : no distorted points;

x_c and y_c : image center.

In figure [Fig. 13], the two curves pass through a minimum. This minimum does not correspond to 0% distortion.

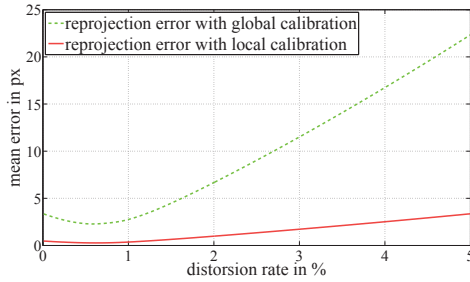


Fig. 13 Simulation for reprojection with local and global calculation. The distortion ranges from 0 to 5 %.

tion but is found at 0.8%. In this position, the optical lens distortion of the scene camera is partially compensated for the distortion due to the spherical shape of the eye. These curves show that the local relationship method increase the precision of POR estimation significantly, where the distortion is important.

5.2 Experimental results

The simulation results are confirmed in an experimental context. Figure [Fig. 14] shows tests carried out through auto-calibration with global and local relationship methods. In this image, lens with a focal length of $3.8mm$ is used, and optical distortion is well visible. The local relationship helps to position the POR in the centre of the target viewed at by the user. This method is well adapted for eye tracker mounted with short focal.

6 Parallax influence and epipolar line computation

Monocular Eye Trackers are designed in a way to estimate the POR based on calibration phase done in a specific distance; which means that accuracy of the POR depends greatly of how far from the calibration depth the gaze is directed. In the experimental conditions, it is impossible to respect this condition exactly. The user can look at road signs or cars, located at any depth. An efficient solution could consist in using a binocular ET [11], but this solution would be too intrusive for learner drivers. Eye Tracker is a very useful tool for school driver, but it can be used routinely only if the calibration is simple. This implies that the user has to calibrate the ET system inside the car and not outside, to avoid the assistance of the driving teacher to carry the calibration pattern outside at the start of a driving lesson. The consequence is that the POR is affected by the parallax, because the user calibrates approximately

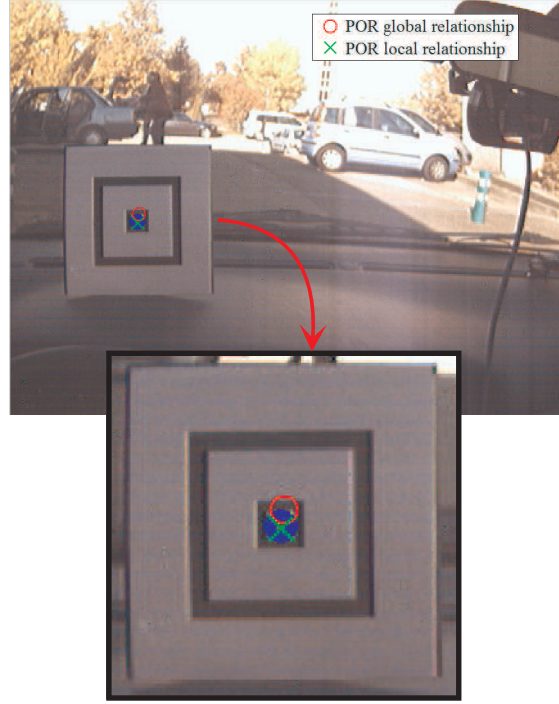


Fig. 14 Experimental results through a calibration with a single point. Local method: green and global method: in red. The distance between the two solutions is 7 pixels.

at 1 metre, and will usually look at more than 5 metres during the whole lesson. The goal of this section is to propose an efficient method to correct the POR, by considering that for our application, we need to analyse the driver gaze position when he looks at outside the vehicle.

In this section, we start to explain that a monocular ET system is a complex stereovision system and can therefore be modeling as such. In the next paragraph, an expression of the parallax error is given, and we deduced that it is possible to rely the epipolar line expression to the parallax error, to avoid to compute a no standard fundamental matrix. In the last paragraph of this section, we present the results, and their interests to compute the POR, when the calibration is done inside the car.

6.1 Stereovision model of ET

Figure [Fig. 15] gives a geometric representation of a monocular ET system. This figure shows that a point P located on the gaze direction is projected at different coordinates in the scene camera following its depth. We can take into consideration that the scene and eye cameras form a stereovision system. Unfortunately, we are

$$p_s^t F' p_e = 0 \quad (15)$$

Where F' is the combination between H and F , assuming, in this example, that H is a homography :

$$F' = HF \quad (16)$$

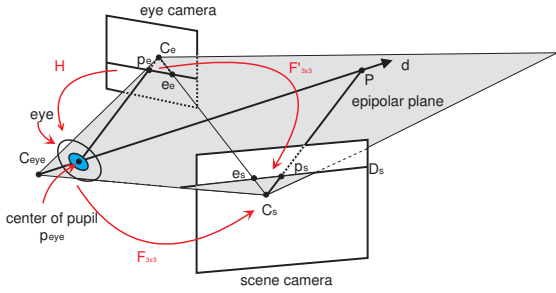


Fig. 15 Model of the ET system and the eye. Point P is in the same projection on the eye and the eye camera anywhere on line D . But the projections from point P in the scene camera are not merged, the projections form an epipolar line D_s

6.2 Calibration distance and parallax error

This paper proposes an alternative based on the calculation of parallax. The reprojection error of the POR estimation is directly related to the distance calibration. [7] describes a close approach by considering the

eye and the scene camera in the same plane and gives parallax error results in degrees. Here this calculation is extended, using intrinsic and projective matrices to calculate the error of POR in pixels, with the geometrical model presented in [Fig. 16].

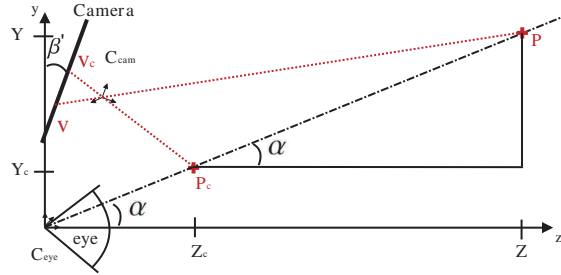


Fig. 16 Parallax between the eye and the scene camera. Points P_c and P are the same projection on the eye in two positions, but the projections of the point P_c and P in the scene camera are not merged. Z_c is the calibration distance.

$$\underbrace{s \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}}_p = \underbrace{\begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & -\alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{intrinsic matrix}} \underbrace{\begin{pmatrix} 1 & 0 & 0 & T_x \\ 0 & \cos(\beta) & -\sin(\beta) & T_y \\ 0 & \sin(\beta) & \cos(\beta) & T_z \end{pmatrix}}_{\substack{\text{rotations and} \\ \text{translations matrix}}} \underbrace{\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}}_P \quad (17)$$

Given a 3D point P_c , inside the calibration plane (ie $Z = Z_c$) its projection is expressed as (from equation 17):

$$u_c = \frac{\alpha_u X_c + u_0 \sin(\beta) Y_c}{Y_c \sin(\beta) + Z_c \cos(\beta) + T_z} + \frac{u_0 \cos(\beta) Z_c + \alpha_u T_x + u_0 T_z}{Y_c \sin(\beta) + Z_c \cos(\beta) + T_z} \quad (18)$$

$$v_c = \frac{(-\alpha_v \cos(\beta) + v_0 \sin(\beta))Y_c}{Y_c \sin(\beta) + Z_c \cos(\beta) + T_z} + \frac{(-v_0 \cos(\beta) - \alpha_v \sin(\beta))Z_c - \alpha_v T_y + v_0 T_z}{Y_c \sin(\beta) + Z_c \cos(\beta) + T_z} \quad (19)$$

Idem for P located at depth Z on visual axes:

$$u(Z) = \frac{\alpha_u X + u_0 \sin(\beta) Y}{Y \sin(\beta) + Z \cos(\beta) + T_z} + \frac{u_0 \cos(\beta) Z + \alpha_u T_x + u_0 T_z}{Y \sin(\beta) + Z \cos(\beta) + T_z} \quad (20)$$

$$v(Z) = \frac{(-\alpha_v \cos(\beta) + v_0 \sin(\beta)) Y}{Y \sin(\beta) + Z \cos(\beta) + T_z} + \frac{(-v_0 \cos(\beta) - \alpha_v \sin(\beta)) Z - \alpha_v T_y + v_0 T_z}{Y \sin(\beta) + Z \cos(\beta) + T_z} \quad (21)$$

The parallax error can be deduced as the following differences:

$$d_u(Z) = u(Z) - u_c = \frac{(\alpha_u T_x + u_0 T_z) \left(\frac{Z_c - Z}{Z_c Z} \right)}{\tan(\alpha) \sin(\beta) + \cos(\beta)} \quad (22)$$

$$d_v(Z) = v(Z) - v_c = \frac{(-\alpha_v T_y + v_0 T_z) \left(\frac{Z_c - Z}{Z_c Z} \right)}{\tan(\alpha) \sin(\beta) + \cos(\beta)} \quad (23)$$

Expressions 22 and 23 are obtained by cancelling $\frac{T_z}{Z_c}$ and $\frac{T_z}{Z}$ compared to others denominators terms. These expressions allow to compute $d_u(Z)$ and $d_v(Z)$ in function of Z . All parameters are known, excepted $\tan(\alpha)$. $\tan(\alpha)$ can be determined by solving equation (19), assuming that $\tan(\alpha) = Y_c/Z_c$. The term v_c of equation (19) is the vertical coordinates of the POR, and it's computed before by mapping with the local relationship.

[Fig. 17] presents parallax error simulations according to several calibration distances. Figure [Fig. 18] shows the superposition of simulation and experimental measurements when calibration is performed at 2.5m. In this figure the variations of the experimental results are caused by the difficulties of displacing an object in the optical axis of a user. Theses curves demonstrate that there is no optimal calibration distance. With a calibration at 5 metres, where the calibration pattern is outside the car, the error on POR lies under 10 pixels for objects seen at more than 5 metres. This error is dramatically important if the user looks at objects at less than 5 metres.

For the present application, the ET system is calibrated at approximately 1 metre. In this case [Fig. 17] the error on POR can reach 40 pixels. This is a very important error, but the advantage is that the calibration process can be performed in the car, which is more comfortable for the learner and his instructor. The next paragraph will show that the previous relation expressing the parallax error can be used to correct the POR position, and deduce the position of the corresponding epipolar line.

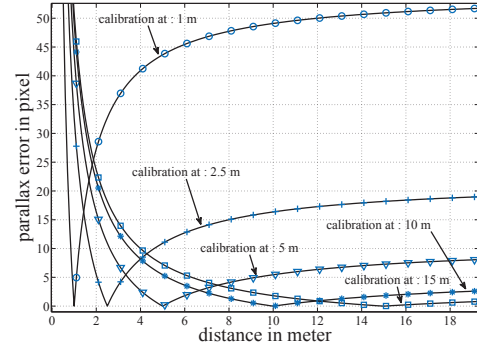


Fig. 17 Miscalculation of the POR due to the parallax phenomenon over a function of the distance.

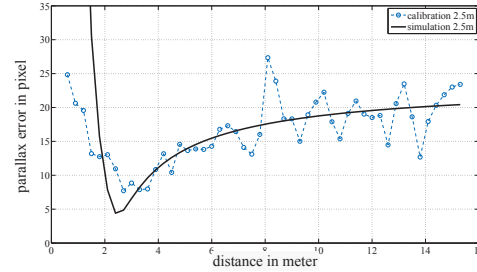


Fig. 18 Comparison between simulation and real measurements, for calibration done at 2.5 metres.

6.3 Compute epipolar line from parallax expressions

Equations (22) and (23) give parallax error of our head mounted ET system. These equations traduce the variation of point of regard on the epipolar line. This allows to compute the POR if the user is looking at any depth, by adding these variations to the POR computed at the distance of calibration Z_c with the local relationship. If the user is looking an object at a depth Z , the POR (Z) is given by equations:

$$u_{POR}(Z) = u_{POR} + d_u(Z) \quad (24)$$

$$v_{POR}(Z) = v_{POR} + d_v(Z) \quad (25)$$

The position of the scene camera can be different following the user's morphology, but tests carried out with more than 20 people show that these positions can be set at $T_x = -3.5$ cm, $T_y = 5.5$ cm, $T_z = 2$ cm. Even if the translations are not well determined, the incertitude of POR can be reduced to a quarter of a disk, as will be shown in the next section.

6.4 Experimentals results

Two cases are considered when the ET is calibrated at 1 metre (inside the car). In the first case, the position of the scene camera is not exactly known from the user's eye. In the second case, it is known.

When the exact position of the scene camera is not available, the geometry of the system shows that T_x is negative, T_y and T_z are positive, and so, the ratio $d_v(Z)/d_u(Z)$ is thus positive. It can be deduced that the real POR position should be in a quarter of a disk of radius 50 pixels. This radius corresponds to the maximal amplitude of the parallax error obtained from the curves in [Fig. 17]. Images of [Fig. 19] illustrate this case. First, the POR is computed with calibration at one metre when the user looks at a road sign at 10 metres approximately; the parallax produces a wrong position for the POR. In second step, the quarter disk is superposed on the image and the true POR (ie road sign) belongs to the area covered by this quarter disk. This reduces the parallax influence and indicates that the user has look at inside this disk quarter. It is assumed that the learner has really seen the road sign.



Fig. 19 The driver looks at a road sign. The ET system is calibrated at 1m and the object viewed is at 10m approximately. The uncertainty range of the parallax phenomenon is the quarter of the blue disk.

In the case where camera position is available, it is possible to compute the epipolar line from equations (24) and (25). This has been done on figure [Fig. 20], and the parametric formulation of corresponding epipolar line allows to draw epipolar line from $Z = 0.1m$ to infinity. This information is interesting, but can't be usable in this form. However, the parametric formulation of the epipolar line has still some interests since the POR can

be computed as a function of the object depth. An interesting compromise, when ET is calibrated inside the car, consists in reducing the depth between 6 metres and infinity. The corresponding portion of the epipolar line is then a small segment of 15 pixels approximately. An example is given on figure [Fig. 21], and we can show that residual incertitude is reduced, and the remaining segment is a reliable marker to represent the POR. It is valid, of course, only under the assumption the ET is used to study how the driver looks at and analyzes the driving scene outside the vehicle, which is exactly the case for our application.

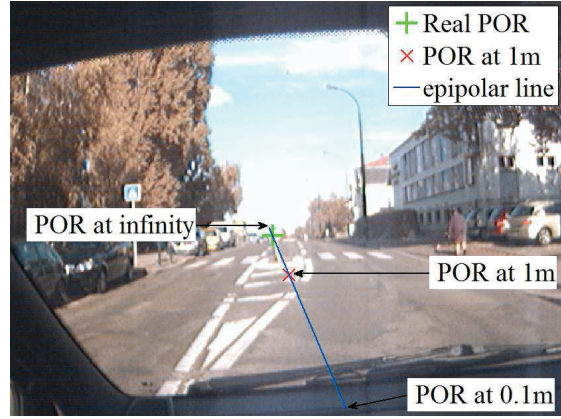


Fig. 20 The driver looks at a road sign. The ET system is calibrated at 1m and the object viewed is at 10m approximately. By measuring T_x , T_y and T_z it is possible to reduce the uncertainty range of the parallax phenomenon and thus draw an epipolar line.

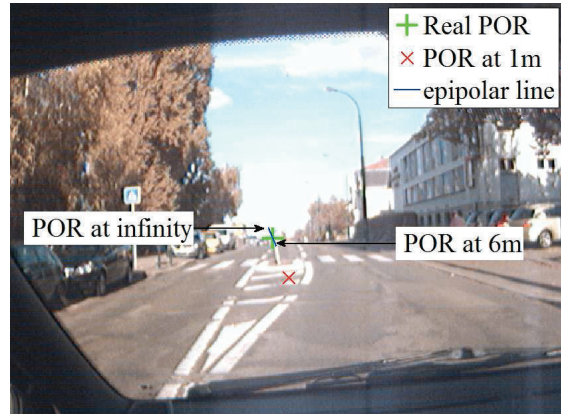


Fig. 21 The POR estimated at 1 metre is substituted by a segment of the epipolar line. The segment starts for $Z = 6m$ and ends for $Z \rightarrow \infty$. This segment mesure approximately 15 pixels.

6.5 Discussion

The proposed method to update the POR, takes advantage of the stereovision model of ET system. The benefit of the easy calibration procedure presented previously are preserved, since the POR is first calculated through mapping, then updated according to the position of the scene camera. Concerning the repeatability of this process, we can deduce from equations (22) and (23), that the slope of the epipolar line, given by $d_v(Z)/d_u(Z)$, is Z free and constant. That means that epipolar lines are parallel in scene image. If we analyze the geometric configuration between eye's user and scene camera [Fig. 16], this is an expected result. The interest to work with $d_u(Z)$ and $d_v(Z)$, instead of the slope directly, is to keep the depth Z to compute the length of the marker representing the POR position in function of Z . In consequence, the repeatability depends essentially of the calibration quality or pupil detection, used to compute the POR. For the present application, where calibration must be as simple as possible, the calibration phase is performed in the car, and this solution facilitates its implementation for driving lessons. This system is currently ready to be implemented in driving schools, and experimental videos obtained confirm the interest of parallax correction.

7 Conclusion

This paper has presented a fast method for the auto-calibration of an eye tracker system for inexperienced ET users. The calibration obtained with a global relationship is as precise as standard calibration methods using a grid of points. The advantage is that the calibration procedure is entirely automatic and only lasts a few seconds. Auto-calibration can produce a large set of correspondences which allows the computation of the POR with local relationships. This method is particularly efficient when a short focal lens is mounted on the scene camera. As shown previously, the best POR location can be extracted while the ET system is used for distance different than the calibration one. Results, in this paper, were obtained by modelling the ET as an indirect stereovision system. New ET approach helps to retrieve the epipolar line in the scene image and gives a reliable guess of the real POR position when the user looks anywhere. For automotive applications, this approach allows fast 2D calibration by placing only one target in the car.

It may be possible to extend this stereovision model to a binocular system. In this case, epipolar crossing

should give a precise POR position and 3D position of the object seen by the user can be deduced.

Acknowledgements The authors wish to thank Joel Lambert for the design and creation of ET prototypes.

References

1. Basset, M., Cudel, C., Gissinger, G.: Visual characterization of the road driver's behavior. In: *Intelligent Signal Processing* (2005)
2. Bernet, S., Sturm, P., Cudel, C., Basset, M.: Study on the interest of hybrid fundamental matrix for head mounted eye tracker modeling. In: *Proceedings of the British Machine Vision Conference*, pp. 15.1–15.10. BMVA Press (2011)
3. Böhme, M., Dorr, M., Graw, M., Martinetz, T., Barth, E.: A software framework for simulating eye trackers. In: *ETRA '08: Proceedings of the 2008 symposium on Eye tracking research and applications*, pp. 251–258. ACM, New York, NY, USA (2008)
4. Crow, F.C.: Summed-area tables for texture mapping. In: *SIGGRAPH '84: Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, pp. 207–212. ACM, New York, NY, USA (1984)
5. Cudel, C., Bernet, S., Basset, M.: Fast and easy calibration for a head mounted eye tracker. In: *COGAIN Communication, Environment and Mobility Control by Gaze*, pp. 11–14 (2008)
6. Djamasbi, S., Siegel, M., Tullis, T.: Generation y, web design, and eye tracking. *International Journal of Human-Computer Studies* **68**(5), 307 – 323 (2010)
7. Dongheng, L.: Low-cost eye-tracking for human computer interaction. Master's thesis, Iowa State University (2006)
8. Dongheng, L., David, W., Derrick, J.P.: Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. *Computer Vision and Pattern Recognition Workshop* **0**, 79 (2005)
9. Droege, D., Schmidt, C., Paulus, D.: A comparison of pupil centre estimation algorithms. In: *4th Conference on Communication by Gaze Interaction COGAIN 2008*. Prague, Czech Republic (2008)
10. Duchowski, A.T.: *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2007)
11. Duchowski, A.T., Medlin, E., Gramopadhye, A., Melloy, B., Nair, S.: Binocular eye tracking in vr for visual inspection training. In: *VRST '01: Proceedings of the ACM symposium on Virtual reality software and technology*, pp. 1–8. ACM, New York, NY, USA (2001)
12. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* **24**(6), 381–395 (1981)
13. Franchak, J.M., Adolph, K.E.: Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research* **50**(24), 2766 – 2774 (2010). *Perception and Action: Part I*
14. Gruen, A., Huang, T.S.: *Calibration and Orientation of Cameras in Computer Vision*. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2001)
15. Hansen, D.W., Agustin, J.S., Villanueva, A.: Homography normalization for robust gaze estimation in uncalibrated setups. In: *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications, ETRA '10*, pp. 13–20. ACM, New York, NY, USA (2010)

16. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, New York, NY, USA (2003)
17. Howell, I., Aulikki, H., Stephen, V., Ali, N.: User performance of gaze-based interaction with on-line virtual communities. In: Communication by Gaze Interaction (COGAIN), pp. 28 – 32 (2008)
18. ICNIRP: Light-emitting diodes (leds) and laser diodes: Implications for hazard assessment. Health Physics **78**(6), 744–752 (2000)
19. Kumar, N., Kohlbecher, S., Schneider, E.: A novel approach to video-based pupil tracking. In: SMC'09: Proceedings of the 2009 IEEE international conference on Systems, Man and Cybernetics, pp. 1255–1262. IEEE Press, Piscataway, NJ, USA (2009)
20. Muller P. U. and Cavegn, D., d'Ydewalle, G., Groner, R.: comparison of a new limbus tracker, corneal reflection technique, purkinje eye tracking and electro-oculography,. Perception and Cognition, Elsevier Science Publishers pp. 393–401 (1993)
21. Noris, B., Benmachiche, K., G. Billard, A.: Calibration-free eye gaze direction detection with gaussian processes. In: Proceedings of the International Conference on Computer Vision Theory and Applications (2008)
22. Peter, B., Anders, H., Bengt, R., John, E., John, E.: Visual display with illuminators for gaze tracking (2011)
23. Petr, N., Tomas, K., Libor, P., Marcel, F., Olga, S.: AI support for a gaze controlled wheelchair. In: Communication by Gaze Interaction (COGAIN), pp. 19 – 22 (2008)
24. Pomplun, M., Velichkovsky, B., Ritter, H.: An artificial neural network for high precision eye movement tracking. In: B. Nebel, L. Dreschler-Fischer (eds.) KI-94: Advances in Artificial Intelligence, *Lecture Notes in Computer Science*, vol. 861, pp. 63–69. Springer Berlin Heidelberg (1994)
25. Putra, D., Cahyawan, A., Perdana, Y.: Low-cost based eye tracking and eye gaze estimation. TELKOMNIKA **9**, 377–386 (2011)
26. Ramanauskas, N., Daunys, G., Dervinis, D.: Investigation of calibration techniques in video based eye tracking system. In: Proceedings of the 11th international conference on Computers Helping People with Special Needs, ICCHP '08, pp. 1208–1215. Springer-Verlag, Berlin, Heidelberg (2008)
27. Rash, C.E.: A 25-year retrospective review of visual complaints and illusions associated with a monocular helmet-mounted display. Displays **29**(2), 70 – 80 (2008). Health and Safety Aspects of Visual Displays
28. Scott, D., Findlay, J.M.: Visual search, eye movements and display units, human factors report. Tech. rep., University of Durham (1993)
29. Sliney, D., Wolbarst, M.: Safety with Lasers and Other Optical Sources: A Comprehensive Handbook. Springer, New York, NY, USA (1980)
30. Stefanovic, P.: Relative orientation - a new approach. Tech. rep., Institute for Aerial Survey and Earth Sciences (1973). 418-448
31. Sturm, P., Ramalingam, S.: Gometrie d'images multiples pour des modles de camras gnraux. TDS Traitement du Signal **22**(5), 483–495 (2005)
32. Villanueva, A., Cabeza, R., Porta, S.: Eye tracking: Pupil orientation geometrical modeling. image and vision computing **24**(7), 663–679 (2005)
33. Zhang, Z.: Determining the epipolar geometry and its uncertainty: A review. Int. J. Comput. Vision **27**, 161–195 (1998)